Séminaire Lotharingien de Combinatoire **93B** (2025)
Article #50, 12 pp.

Proceedings of the 37th Conference on Formal Power Series and Algebraic Combinatorics (Sapporo)

Principal minors of tree distance matrices (extended abstract)

Harry Richman[*1], Farbod Shokrieh[2], and Chenxi Wu[3]

[1]*Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan*
[2]*Department of Mathematics, University of Washington, Seattle, USA*
[3]*Department of Mathematics, University of Wisconsin, Madison, USA*

Abstract. Suppose D is the distance matrix of a tree. Graham and Pollack showed that the determinant of D satisfies a surprising identity that depends only on the number of vertices in the given tree. We generalize this result to a combinatorial identity for the determinant of any principal submatrix of D. This new identity involves counts of spanning forests and is proved by use of potential-theoretic concepts on graphs.

Keywords: tree, distance matrix, potential theory, spanning forest

1 Introduction

For two vertices u, v in a connected graph G, the *path distance* $d(u, v)$ is the number of edges on a shortest path from u to v. For brevity, we will use "distance" to mean path distance in this article. The *distance matrix* of G is the matrix $D \in \mathbb{R}^{V \times V}$ whose (u, v)-entry is the distance $d(u, v)$.

As a combinatorialist, perhaps inspired by the matrix-tree theorem, one may ask:

Question 1.1. Is there any combinatorial information encoded in the determinant of the distance matrix of a graph?

Example 1.2. Suppose that G is the tree shown below in Figure 1.

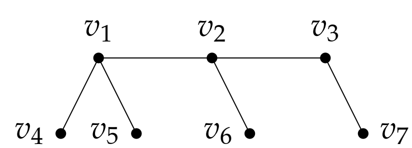

Figure 1: Tree with seven vertices and four leaves.

*hrichman@ncts.ntu.edu.tw. H. R. was partially supported by the Howard Hughes Medical Institute and the Matsen Group at the Fred Hutchinson Cancer Center. F. S. was partially supported by NSF CAREER grant DMS-2044564. C. W. was partially supported by Simons Collaboration Grant 850685.

Using the ordering indicated by the vertex labels, the tree G has distance matrix

$$D = \begin{pmatrix} 0 & 1 & 2 & 1 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 & 2 & 1 & 2 \\ 2 & 1 & 0 & 3 & 3 & 2 & 1 \\ 1 & 2 & 3 & 0 & 2 & 3 & 4 \\ 1 & 2 & 3 & 2 & 0 & 3 & 4 \\ 2 & 1 & 2 & 3 & 3 & 0 & 3 \\ 3 & 2 & 1 & 4 & 4 & 3 & 0 \end{pmatrix}.$$

By direct computation, this matrix has determinant $\det D = 192$. $\diamond$

A result of Graham and Pollak [11] states that if G is a tree with n vertices, then its distance matrix D satisfies

$$\det D = (-1)^{n-1} 2^{n-2} (n-1). \tag{1.1}$$

This remarkable identity says that when G is a tree, essentially no combinatorial information from G is retained in $\det D$. Thus in this case, the answer to Question 1.1 is an emphatic "no"!

1.1 Distance submatrices

If we are optimistic that there is interesting combinatorics hidden somewhere within the distance matrix D, despite the Graham–Pollak identity (1.1), then we could next consider its minors, i.e., determinants of submatrices. For a vertex subset $S \subset V(G)$ of a graph G, let $D[S]$ denote the submatrix consisting of the S-indexed rows and columns of D.

Question 1.3. Is there any combinatorial information encoded in the determinant of the distance *submatrix* $D[S]$ of a graph?

Our main contribution answers "yes" to Question 1.3 when the graph is a tree, in the form of the following theorem.

Theorem 1.4 ([14, Theorem A])**.** *Suppose G is a tree with n vertices, and distance matrix D. Let $S \subset V(G)$ be a nonempty subset of vertices. Then*

$$\det D[S] = (-1)^{|S|-1} 2^{|S|-2} \left((n-1)\, \kappa_1(G;S) - \sum_{\mathcal{F}_2(G;S)} (\deg^o(F,*) - 2)^2 \right), \tag{1.2}$$

where $\kappa_1(G;S)$ is the number of S-rooted spanning forests of G, $\mathcal{F}_2(G;S)$ is the set of $(S,*)$-rooted spanning forests of G, and $\deg^o(F,*)$ is the outdegree of the floating component of F.

Here, an *S-rooted spanning forest* of G means a spanning forest in which each connected component has exactly one vertex of S. An *$(S,*)$-rooted spanning forest* of G is a spanning forest which has $|S|+1$ components, where $|S|$ components each contain one vertex of S, and the additional component is disjoint from S. We call the component disjoint from S the *floating component*, following terminology in [13]. We let $F(*)$ denote the floating component of an $(S,*)$-rooted spanning forest. If G is the tree in Figure 1 and S is the set of four leaf vertices, then the S-rooted spanning forests are shown in Figure 2, and the $(S,*)$-rooted spanning forests are shown in Figure 3.

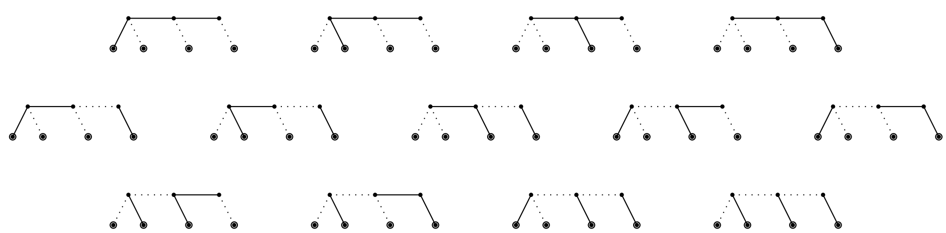

Figure 2: All forests in $\mathcal{F}_1(G;S)$, with vertices in S bolded.

Given a connected subgraph $H \subset G$, the *outdegree* $\deg^o(H)$ of H is the number of edges which have one endpoint in H and the other endpoint outside H. (The outdegree does not depend on edge orientation.) In Figure 3, the forests are grouped according to the outdegree of the floating component; the floating component may have outdegree two (top row), outdegree three (lower left), or outdegree four (lower right).

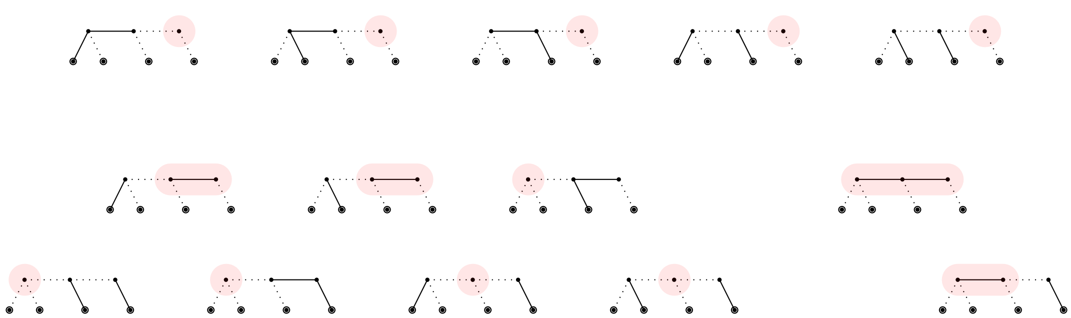

Figure 3: All forests in $\mathcal{F}_2(G;S)$, with floating component highlighted red. The outdegree $\deg^o(F(*))$ is either 2, 3, or 4.

Example 1.5. If G is the tree in Figure 1, and $S=\{v_4,v_5,v_6,v_7\}$ is the set of leaf vertices, then the corresponding distance submatrix is

$$D[S]=\begin{pmatrix}0&2&3&4\\2&0&3&4\\3&3&0&3\\4&4&3&0\end{pmatrix}. \tag{1.3}$$

In this case, the submatrix has determinant $\det D[S] = -252$.

The tree G has 6 edges and 13 S-rooted spanning forests. There are 14 $(S,*)$-rooted spanning forests; of the floating components in these forests, 5 have outdegree two, 7 have outdegree three, and 2 have outdegree four. See Figures 2 and 3. One checks that

$$\det D[S] = -252 = (-1)^3 2^2 \left(6 \cdot 13 - (5 \cdot 0^1 + 7 \cdot 1^2 + 2 \cdot 2^2)\right),$$

in agreement with Theorem 1.4. ◇

We omit the full proof of Theorem 1.4 in this extended abstract, deferring it to our paper [14], but provide the main steps of the argument in Section 2. In Section 1.2 we mention some subsequent results related to Theorem 1.4.

Remark 1.6. When $S = V$ is the full vertex set, the set of V-rooted spanning forests is a singleton, consisting of the subgraph with no edges, so $\kappa_1(G;V) = 1$; and moreover the set $\mathcal{F}_2(G;V)$ is empty. Thus (1.2) recovers the Graham–Pollak identity (1.1) when $S = V$.

Remark 1.7. It is worth observing that depending on the chosen subset $S \subset V$, the distances appearing in the submatrix $D[S]$ may ignore a large part of the ambient tree G. To apply formula (1.2) "efficiently," we could replace G on the right-hand side with the subtree $\operatorname{conv}(S,G)$ consisting of all edges on paths between vertices in S. However, the formulas as stated are true even without this replacement due to cancellation of terms.

1.2 Cofactor sums

In [5], Bapat and Sivasubramanian answered "yes" to a slight modification of Question 1.3 on trees, without using determinants. They showed that the *sum of cofactors* of a distance submatrix $D[S]$ of a tree satisfies the following identity [5, Theorem 9]:

$$\operatorname{cof} D[S] = (-2)^{|S|-1} \kappa_1(G;S). \tag{1.4}$$

Recall that $\kappa_1(G;S)$ is the number of S-rooted spanning forests of G. Given a matrix A, the sum of cofactors of A is defined as

$$\operatorname{cof} A = \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} (-1)^{i+j} \det A_{i,j}$$

where $A_{i,j}$ is the submatrix of A that removes the i-th row and the j-th column. If A is invertible, then $\operatorname{cof} A$ is the sum of entries of the matrix inverse A^{-1} multiplied by a factor of $\det A$, i.e. $\operatorname{cof} A = (\det A)(\mathbf{1}^\intercal A^{-1} \mathbf{1})$. Here $\mathbf{1}$ denotes the all-ones vector.

Using the Bapat–Sivasubramanian identity (1.4), Theorem 1.4 immediately yields:

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \frac{1}{2} \left((n-1) - \frac{\sum_{F \in \mathcal{F}_2(G;S)} (\deg^o(F,*) - 2)^2}{\kappa_1(G;S)} \right). \tag{1.5}$$

Surprisingly, the expression (1.5) satisfies a monotonicity condition as we fix a tree G and vary the vertex subset $S \subset V(G)$.

Theorem 1.8 (Monotonicity of "normalized" principal minors, [14, Theorem C])**.** *If $A, B \subset V(G)$ are nonempty vertex subsets with $A \subset B$, then*

$$\frac{\det D[A]}{\operatorname{cof} D[A]} \leq \frac{\det D[B]}{\operatorname{cof} D[B]}.$$

The monotonicity result implies the following bounds as a corollary.

Theorem 1.9 (Bounds on "normalized" principal minors, [14, Theorem D])**.** *Suppose G is a tree with distance matrix D.*

(a) If $\operatorname{conv}(S, G)$ *denotes the subtree of G consisting of all edges on paths between vertices of $S \subset V(G)$, then* $\dfrac{\det D[S]}{\operatorname{cof} D[S]} \leq \dfrac{1}{2}|E(\operatorname{conv}(S, G))|$.

(b) If γ is a simple path between vertices $s_0, s_1 \in S$, then $\dfrac{1}{2}|E(\gamma)| \leq \dfrac{\det D[S]}{\operatorname{cof} D[S]}$.

We note that Theorem 1.9 (b) is an equality if S consists of exactly two points. The result in Theorem 1.8 was also independently obtained by Devriendt [8, Property 3.38] in a more general context studying effective resistance on graphs.

1.3 Related work

The initial work of Graham and Pollack [11] inspired a large amount of subsequent research on distance matrices and generalizations. In [9], it is observed that the distance matrix ratio $\det D / \operatorname{cof} D$ is additive over wedge sums of graphs. Bapat, Lal, and Pati [4] defined a q-analogue of the distance matrix and found formulas for its determinant and inverse. Bapat [2] also found a formula for the determinant of the effective resistance matrix of a graph, although the combinatorial content here is more obscured. It would be worth studying whether these results can be extended to arbitrary principal minors.

Choudhury and Khare [6] generalize the Graham–Pollak identity to some principal minors $\det D[S]$ for a certain restricted class of vertex subsets S. They do so using a framework that encompasses the q-analogue distance matrix as well. Very recently, Gutiérrez and Lillo [12] observe that Theorem 1.4 can be expressed as

$$\begin{aligned}\det D[S] = (-1)^{|S|-1}2^{|S|-2}\Big(&(|S|-1)\,\kappa_1(G;S) \\ &- \sum_{F \in \mathcal{F}_2(G;S)} (\deg^o(F,*)-1)\,(\deg^o(F,*)-4)\Big)\end{aligned}$$

through some straightforward algebraic manipulation, while also providing a new proof for this result using a nice combinatorial argument involving sign-reversing involutions on collections of paths in G.

2 Proof roadmap

Here we give a brief outline to how we obtain the expression of $\det D[S]$ that appears in Theorem 1.4, our main result. The argument consists of four steps. (Recall that $\mathbf{1}$ denotes the all-ones vector.)

(I) Find a vector $\mathbf{m} \in \mathbb{R}^S$ such that $D[S]\mathbf{m} = \lambda\mathbf{1}$, for some scalar λ. Find λ.

(II) Compute $\mathbf{1}^\intercal\mathbf{m}$. Since $D[S]$ is nonsingular (cf. Lemma 4.2), we have $\mathbf{m} = \lambda(D[S]^{-1}\mathbf{1})$ and hence

$$\mathbf{1}^\intercal\mathbf{m} = \lambda(\mathbf{1}^\intercal D[S]^{-1}\mathbf{1}) = \lambda \cdot \frac{\operatorname{cof} D[S]}{\det D[S]}.$$

(III) Solve or simplify $\dfrac{\det D[S]}{\operatorname{cof} D[S]} = \dfrac{\lambda}{\mathbf{1}^\intercal\mathbf{m}}$.

(IV) Multiply previous expression by $\operatorname{cof} D[S]$, using (1.4), to get $\det D[S]$.

We demonstrate the steps of this proof roadmap on our running example.

Example 2.1. Suppose G is the tree from Figure 1, which we reproduce in Figure 4. As before, let S consist of the four leaf vertices, which are marked with larger circles.

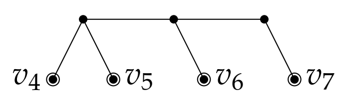

Figure 4: A tree with six edges, and $S \subset V$ containing four vertices.

The distance submatrix is $D[S] = \begin{pmatrix} 0 & 2 & 3 & 4 \\ 2 & 0 & 3 & 4 \\ 3 & 3 & 0 & 3 \\ 4 & 4 & 3 & 0 \end{pmatrix}$. We now compute $\det D[S]$ via the above roadmap.

(I) The vector $\mathbf{m} = \begin{pmatrix} 6 & 6 & 5 & 9 \end{pmatrix}^\intercal$ satisfies $D[S]\mathbf{m} = \lambda\mathbf{1}$ for $\lambda = 63$.

(II) We have $\mathbf{1}^\intercal\mathbf{m} = 26$.

(III) Thus $\dfrac{\det D[S]}{\operatorname{cof} D[S]} = \dfrac{\lambda}{\mathbf{1}^\intercal\mathbf{m}} = \dfrac{63}{26}$.

(IV) Using (1.4), the cofactor sum is $\operatorname{cof} D[S] = -104$, so

$$\det[S] = (\operatorname{cof} D[S]) \cdot \left(\frac{\det D[S]}{\operatorname{cof} D[S]}\right) = -252. \qquad \diamond$$

While not obvious here, it turns out that the entries of $\mathbf{m}$ are combinatorially meaningful (see Definition 3.1), which also gives combinatorial meaning to the constant λ.

3 Potential theory

Here we discuss step (I) of the roadmap (Section 2), which involves solving $D[S]\mathbf{m} = \lambda\mathbf{1}$ for some $\mathbf{m} \in \mathbb{R}^S$ and real λ. This problem can be viewed through the lens of potential theory: $\mathbf{m}$ describes the distribution of charged particles among the vertices S, and the entries of $D[S]$ give the (negative) potential stored from interacting charges repelling one another in the ambient space G, depending on their position. If the potential $D[S]\mathbf{m}$ is non-constant, charges will move from higher-potential locations to lower-potential ones (leaving total charge conserved) in order to decrease the energy $\mathcal{E}(\mathbf{m}) := \mathbf{m}^\intercal D[S]\mathbf{m}$. The charges will redistribute until they settle in equilibrium, such that $D[S]\mathbf{m}$ is constant on S. The aim of this section is to describe the resulting *equilibrium distribution* $\mathbf{m}$ (up to scaling) and *equilibrium energy* $\mathbf{m}^\intercal D[S]\mathbf{m}$.

We first define a vector $\mathbf{m}$ which satisfies the relation $D[S]\mathbf{m} = \lambda\mathbf{1}$ for some λ.

Definition 3.1. Let $\mathbf{m} = \mathbf{m}(G;S)$ denote the vector in $\mathbb{R}^S$ defined by

$$\mathbf{m}_v = \sum_{T\in\mathcal{F}_1(G;S)} (2 - \deg^o(T,v)) \qquad \text{for each } v \in S, \tag{3.1}$$

where $\deg^o(T,v)$ is the outdegree of the v-component of T.

Proposition 3.2. *The vector* $\mathbf{m} = \mathbf{m}(G;S)$ *defined above satisfies* $\mathbf{1}^\intercal\mathbf{m} = 2\,\kappa_1(G;S)$.

Remark 3.3. The vector $\mathbf{m}$ (3.1) appears in several places in the literature, with various rescalings. In Bapat–Sivasubramanian [5, Equation (20)], it is denoted τ. In Devriendt [8], it is denoted $\mathbf{p}$ called the *resistance curvature* vector of the Kron reduction G/S^c; the corresponding λ, such that $D[S]\mathbf{p} = \lambda\mathbf{1}$, is called the *resistance radius* (up to rescaling by a multiple of $\kappa(G;S)$).

3.1 Main computation

The following is the technical heart of our main result. See [14] for the proof.

Theorem 3.4. *With* $\mathbf{m} = \mathbf{m}(G;S)$ *defined as in* (3.1), $D[S]\mathbf{m} = \lambda\mathbf{1}$ *for the constant*

$$\lambda = (n-1)\kappa_1(G;S) - \sum_{\mathcal{F}_2(G;S)} (2 - \deg^o(F,*))^2. \tag{3.2}$$

Remark 3.5. A key step in the proof of Theorem 3.4 uses the following "transition structure" which relates the S-rooted spanning forests $\mathcal{F}_1(G;S)$ with $(S,*)$-rooted spanning forests $\mathcal{F}_2(G;S)$, via the operations of edge-deletion and edge-union.

Consider the "deletion" map

$$E(G) \times \mathcal{F}_1(G;S) \xrightarrow{\text{del}} \mathcal{F}_1(G;S) \sqcup \mathcal{F}_2(G;S)$$

defined by $\mathrm{del}(e,T) = T \setminus e$ if $e \in T$, and $\mathrm{del}(e,T) = T$ otherwise. For a given spanning forest $F \in \mathcal{F}_2(G;S)$, there are exactly $\deg^o(F,*)$-many choices of pairs $(e,T) \in E(G) \times \mathcal{F}_1(G;S)$ such that $F = T \setminus e$.

There is an associated "union" map

$$E(G) \times \mathcal{F}_2(G;S) \xrightarrow{\mathrm{uni}} \mathcal{F}_1(G;S) \sqcup \mathcal{F}_2(G;S)$$

defined by $\mathrm{uni}(e,F) = F \cup e$ if $e \in \partial F(*)$, and $\mathrm{uni}(e,F) = F$ otherwise. Here, $\partial F(*)$ denotes the set of edges that have one end in the floating component $F(*)$, and one end outside of it. For a spanning forest $T \in \mathcal{F}_1(G;S)$, there are exactly $(|V| - |S|)$-many choices of pairs $(e,F) \in E(G) \times \mathcal{F}_2(G;S)$ such that $T = F \cup e$ (since $|E(T)| = |V| - |S|$ for any such spanning forest T).

4 Monotonicity via quadratic optimization

Here we discuss another rather surprising property of the ratio $\dfrac{\det D[S]}{\operatorname{cof} D[S]}$, which allows us to prove Theorem 1.8. Throughout, we assume that D is the distance matrix of a tree. Recall that $\operatorname{cof} M$ denotes the sum of cofactors of a matrix M.

Proposition 4.1. *We have* $\dfrac{\det D[S]}{\operatorname{cof} D[S]} = \max\{\mathbf{u}^\intercal D[S]\mathbf{u} : \mathbf{u} \in \mathbb{R}^S,\ \mathbf{1}^\intercal \mathbf{u} = 1\}$.

Proposition 4.1 is shown in two steps: first, by use of Lagrange multipliers we can show that $\frac{\det D[S]}{\operatorname{cof} D[S]}$ is a critical value. This part is straightforward. The second step is to confirm that the critical value is indeed a maximum, by analysis of the signature of $D[S]$.

Lemma 4.2 (Bapat [3, Lemma 8.15])**.** *Suppose D is the distance matrix of a tree with n vertices. Then D has one positive eigenvalue and $n-1$ negative eigenvalues.*

Starting from Lemma 4.2, it is straightforward to argue that the submatrix $D[S]$ has signature $(1, |S| - 1)$; see [14, Lemma 3.4].

Proposition 4.1 is essentially enough for us to prove our result on the monotonicity of the ratios $\frac{\det D[S]}{\operatorname{cof} D[S]}$, as S varies over vertex subsets of a fixed tree.

We first give a minor restatement of Proposition 4.1, viewing $\mathbb{R}^S$ as a subspace of $\mathbb{R}^V$ where coordinates indexed by $V \setminus S$ are set to zero.

Corollary 4.3. *We have* $\dfrac{\det D[S]}{\operatorname{cof} D[S]} = \max\{\mathbf{u}^\intercal D\mathbf{u} : \mathbf{u} \in \mathbb{R}^V,\ \mathbf{1}^\intercal \mathbf{u} = 1,\ \mathbf{u}_v = 0 \text{ if } v \notin S\}$.

We now return to Theorem 1.8.

Proof of Theorem 1.8. We are to show that for vertex subsets $A \subset B$, we have $\frac{\det D[A]}{\operatorname{cof} D[A]} \leq \frac{\det D[B]}{\operatorname{cof} D[B]}$. By Corollary 4.3, both values $\frac{\det D[A]}{\operatorname{cof} D[A]}$ and $\frac{\det D[B]}{\operatorname{cof} D[B]}$ arise from optimizing the same objective function on an affine subspace of $\mathbb{R}^V$, but the subspace for A is contained in the subspace for B. □

5 Extensions and applications

5.1 Weighted graphs

If a tree $G = (V, E)$ is assigned positive real edge weights $\{\alpha_e : e \in E\}$, then the weighted path distance $d^{(\alpha)}(u, v)$ is defined as the sum of the weights of edges on the unique path from u to v. Bapat–Kirkland–Neumann [1] proved an analogue of (1.1) for the weighted distance matrix of a tree,

$$\det D = (-1)^{n-1} 2^{n-2} \sum_{e \in E} \alpha_e \prod_{e \in E} \alpha_e. \tag{5.1}$$

We additionally have the following weighted version of our main result, which reduces to Theorem 1.4 when taking all unit weights, $\alpha_e = 1$. We defer the proof to the paper [14].

Theorem 5.1. *Suppose $G = (V, E)$ is a tree with distance matrix D. For any nonempty subset $S \subset V$, we have*

$$\det D[S] = (-1)^{|S|-1} 2^{|S|-2} \left(\sum_{E(G)} \alpha_e \sum_{\mathcal{F}_1(G;S)} w(\overline{T}) - \sum_{\mathcal{F}_2(G;S)} w(\overline{F}) \left(\deg^o(F, *) - 2\right)^2 \right), \tag{5.2}$$

where $\mathcal{F}_1(G; S)$ is the set of S-rooted spanning forests T of G, $\mathcal{F}_2(G; S)$ is the set of $(S, *)$-rooted spanning forests F of G, $w(\overline{T})$ and $w(\overline{F})$ denote the co-weights of T and F, and $\deg^o(F, *)$ is the outdegree of the floating component of F, as above.

5.2 Symanzik polynomials

The identity in Theorem 5.1 relates closely to Symanzik polynomials, which are used in quantum field theory for studying Feynman diagrams. We recall the definition here. Given a graph $G = (V, E)$, the *first Symanzik polynomial* is the homogeneous polynomial in edge-indexed variables $\underline{x} = \{x_e : e \in E\}$ defined by

$$\psi_G(\underline{x}) = \sum_{T \in \mathcal{F}_1(G)} \prod_{e \notin T} x_e,$$

where $\mathcal{F}_1(G)$ denotes the set of spanning trees of G.

Consider a "momentum" function $p : V \to \mathbb{R}$ with the constraint $\sum_{v\in V} p(v) = 0$. Then the second Symanzik polynomial is

$$\varphi_G(p;\underline{x}) = \sum_{F\in\mathcal{F}_2(G)} \Big(\sum_{v\in F_1} p(v)\Big)^2 \prod_{e\notin F} x_e,$$

where $\mathcal{F}_2(G)$ is the set of two-component spanning forests of G, and F_1 denotes one of the components[1] of F.

In terms of Symanzik polynomials, let ψ and φ denote the first and second Symanzik polynomials of the quotient graph G/S (i.e., the graph with the same edge set as G, but all vertices in S are glued together to a single vertex). Let p_{can} be the momentum function $p_{can}(v) = \deg(v) - 2$ for $v \notin S$. We have

$$\det D[S] = (-1)^{|S|-1} 2^{|S|-2} \left(\Big(\sum_{E(G)} \alpha_e\Big) \psi_{(G/S)}(\underline{\alpha}) - \phi_{(G/S)}(p_{can};\underline{\alpha}) \right) \tag{5.3}$$

(equivalent to Theorem 5.1), or more succinctly,

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \frac{1}{2}\left(\sum_{e\in E} \alpha_e - \frac{\varphi_{(G/S)}(p_{can};\underline{\alpha})}{\psi_{(G/S)}(\underline{\alpha})} \right) \tag{5.4}$$

(equivalent to a weighted version of equation (1.5)).

5.3 Phylogenetics

Theorem 1.9 may be of interest to those studying phylogenetics. In phylogenetics, one aims to find the tree that best represents the evolutionary history among a collection of organisms using biological data (e.g. DNA sequences). In this tree, leaf vertices represent modern-day species, while internal vertices represent ancestral species. There are standard methods for estimating pairwise distances between species along their evolutionary tree. This means we can often predict the distance submatrix $D[S]$ of the target tree, in which S is the set of leaf vertices, and we would like to use this information in reverse to decide what underlying tree best fits this distance data. Results such as Theorem 1.8 may lead to new tests for phylogenetic inference, or for evaluating tree instability [7].

[1] It doesn't matter which component we label as F_1, since the momemtum constraint implies that $\sum_{v\in F_1} p(v) = -\sum_{v\in F_2} p(v)$.

5.4 Characteristic polynomial coefficients

In [10], Graham and Lovasz study the characteristic polynomial of the distance matrix of a tree $\det(D - \lambda I) = \sum_{k \geq 0} \delta_k \lambda^k$. Note that $\delta_0 = \det D$. They generalize the determinant formula (1.1) by finding a combinatorial formula for every coefficient $\delta_k = \delta_k(G)$ of the characteristic polynomial. Their expression for δ_k involves summing over all spanning forests of G with $(k-1)$-, k-, or $(k+1)$-many edges.

The simplest case of their theorem, beyond δ_0, is [10, Equation (6), p. 63]

$$\delta_1(G) = (-1)^{n-1} 2^{n-3} \left(4 \cdot N_{2P_1}(G) + 2 \cdot N_{P_2}(G) + 4n - 8\right) \tag{5.5}$$

where G is a tree on n vertices, and $N_{2P_1}(G)$ (resp. $N_{P_2}(G)$) denotes the number of subgraphs of G isomorphic to two disjoint edges (resp. to a two-edge path).

There is a well-known relationship between coefficients of the characteristic polynomial and principal minors, namely

$$\delta_k = (-1)^k \sum_{|S|=n-k} \det D[S]. \tag{5.6}$$

By applying Theorem 1.4 and summing over vertex subsets of fixed size, we immediately obtain another combinatorial expression for δ_k. Interestingly, this expression for δ_k looks rather different from [10, Equation (42), p. 81] by Graham and Lovasz, as it involves summing over spanning forests of G with $(n-k)$- or $(n-k+1)$-many components.

It may be possible to give an alternative proof of the Graham–Lovasz identities using Theorem 1.4, which we leave for future work. For now we only elaborate on the $k = 1$ case: if $S = V \setminus \{v\}$ contains $n-1$ vertices, then

$$\kappa_1(G; S) = \deg(v), \qquad \kappa_2(G; S) = 1, \quad \text{and} \quad \deg^o(F, *) = \deg(v)$$

for the unique forest $F \in \mathcal{F}_2(G; S)$. Summing over all vertex subsets with $|S| = n-1$, we obtain by (5.6) and (1.2) that

$$\delta_1(G) = (-1)^{n-1} 2^{n-3} \left((n-1) \sum_{v \in V} \deg(v) - \sum_{v \in V} (\deg(v) - 2)^2 \right), \tag{5.7}$$

which further simplifies to $\delta_1(G) = (-1)^{n-1} 2^{n-3} \left(2(n-1)^2 - \sum_{v \in V} (\deg(v) - 2)^2\right)$.

Acknowledgements

The authors would like to thank Ravindra Bapat for helpful discussion, in particular for providing a pointer to Lemma 4.2. We thank Erick Matsen and the members of his lab for helpful feedback on this work. We also thank Matt Baker, Karel Devriendt, Apoorva Khare, and Sebastian Prillo for helpful discussions. We thank Álvaro Gutiérrez for helpful discussions regarding his follow-up to our work. We thank the anonymous reviewers for feedback that improved the exposition in numerous places.

References

[1] R. Bapat, S. J. Kirkland, and M. Neumann. "On distance matrices and Laplacians". *Linear Algebra Appl.* **401** (2005), pp. 193–209. DOI.

[2] R. B. Bapat. "Resistance matrix of a weighted graph". *MATCH Commun. Math. Comput. Chem.* 50 (2004), pp. 73–82.

[3] R. B. Bapat. *Graphs and matrices*. Universitext. Springer, London; Hindustan Book Agency, New Delhi, 2010, pp. x+171. DOI.

[4] R. B. Bapat, A. K. Lal, and S. Pati. "A q-analogue of the distance matrix of a tree". *Linear Algebra Appl.* **416**.2-3 (2006), pp. 799–814. DOI.

[5] R. B. Bapat and S. Sivasubramanian. "Identities for minors of the Laplacian, resistance and distance matrices". *Linear Algebra Appl.* **435**.6 (2011), pp. 1479–1489. DOI.

[6] P. N. Choudhury and A. Khare. "Distance matrices of a tree: two more invariants, and in a unified framework". *Eur. J. Comb.* **115** (2024). Id/No 103787, 30 pp. DOI.

[7] L. Collienne, M. Barker, M. A. Suchard, and F. A. Matsen IV. "Phylogenetic tree instability after taxon addition: empirical frequency, predictability, and consequences for online inference". *Systematic Biology* (Oct. 2024), syae059. DOI.

[8] K. Devriendt. "Graph geometry from effective resistances". Ph.D. Thesis. 2022.

[9] R. L. Graham, A. J. Hoffman, and H. Hosoya. "On the distance matrix of a directed graph". *J. Graph Theory* **1**.1 (1977), pp. 85–88. DOI.

[10] R. L. Graham and L. Lovász. "Distance matrix polynomials of trees". *Adv. in Math.* **29**.1 (1978), pp. 60–88. DOI.

[11] R. L. Graham and H. O. Pollak. "On the addressing problem for loop switching". *Bell System Tech. J.* **50** (1971), pp. 2495–2519. DOI.

[12] A. Gutiérrez and A. Lillo. "Principal minors of the distance matrix of a tree". 2024. arXiv:2407.01966.

[13] A. Kassel, R. Kenyon, and W. Wu. "Random two-component spanning forests". *Ann. Inst. Henri Poincaré Probab. Stat.* **51**.4 (2015), pp. 1457–1464. DOI.

[14] D. H. Richman, F. Shokrieh, and C. Wu. "Principal minors of tree distance matrices". 2024. arXiv:2411.11488.